



University of Pennsylvania Working Papers in Linguistics

Volume 16

Issue 2 *Selected Papers from NWAV 38*

Article 12

1-1-2010

Do Speech Evaluation Scales in a Speaker Evaluation Experiment Trigger Conscious or Unconscious Attitudes?

Stefan Grondelaers

Radboud University of Nijmegen, S.Grondelaers@let.ru.nl

Roeland van Hout

Radboud University of Nijmegen, R.v.hout@let.ru.nl

This paper is posted at ScholarlyCommons. <http://repository.upenn.edu/pwpl/vol16/iss2/12>

For more information, please contact libraryrepository@pobox.upenn.edu.

Do Speech Evaluation Scales in a Speaker Evaluation Experiment Trigger Conscious or Unconscious Attitudes?

Abstract

This paper challenges the exclusive reliance on speaker personality traits in experimental language attitude research. The most explicitly articulated motivation for this restriction is Kristiansen's (2009) contention that the use of scales pertaining to the investigated varieties themselves could make listener-judges aware of the research purpose, as a result of which their attitudes would be consciously offered, shallow perceptions instead of deeper conceptualizations. On the basis of a free response experiment (in which respondents were invited to explicitly articulate their attitudes towards labelled varieties of Netherlandic Dutch), we demonstrate first that language attitudes are made up of speaker personality traits, but also, and unmistakably, of speech-related perceptions of the investigated varieties. A follow-up experiment (with scaled responses to unlabelled speech stimuli) confirms that the inclusion of speech traits in a speaker evaluation experiment does not affect the nature and structure of the attitudes observed. Building on these data, we argue against the idea that specific attitude measurement techniques should correlate with specific attitude consciousness levels, and we make a plea for a multi-modal approach to language attitudes.

Do Speech Evaluation Scales in a Speaker Evaluation Experiment Trigger Conscious or Unconscious Attitudes?

Stefan Grondelaers and Roeland van Hout*

1 Introduction

A widely accepted but rarely articulated assumption in language attitude research is the idea that attitudes investigated with the Speaker Evaluation paradigm (Lambert et al. 1960) represent “deeper” perceptions than those collected with other techniques. The main reason for this supposedly greater depth is the fact that in Speaker Evaluation (SE) experiments, “respondents have the attitude object (a language, a variety, or even a feature of a variety) presented to them *indirectly*, triggering subconscious evaluation of the linguistic element (the attitude object) under the guise of being asked for an evaluation of the speaker, not his or her linguistic production” (Preston 2009: 270, italics ours). In direct techniques such as label ranking tasks – which overtly elicit language attitudes building on questions such as “which of these varieties do you like best” (Kristiansen 2009) – or free response tasks (see below), respondents are (more) aware of the attitudinal object, as a result of which they offer perceptions which may be more public, explicit, shallow and/or stereotypical (see Garrett 2005 for an overview of attitude measurement techniques and their (dis)advantages).

The most explicit claims made in this respect can be found in Tore Kristiansen’s (2009) attitudinal work on present-day changes in Danish Standard Language Ideologies. Kristiansen explicitly postulates a “division of labor” between direct and indirect techniques, because the first return attitudes which are consciously offered, while the second return attitudes subconsciously offered. Whereas a direct Label Ranking Task confirmed the education-based point of view on Standard Danish—to the effect that the best language in Denmark is the official, conservative standard—indirect SE experiments increasingly reveal that young Danes award the highest prestige to modern Copenhagen speech. This leads Kristiansen to propose that it is the *subconscious* evaluations which are the more dynamic structures that reflect language change.

This paper focuses on a somewhat less audacious, but equally far-reaching claim for SE-based language attitude research in Kristiansen 2009. On page 176, while discussing the methodology of his SE experiment, Kristiansen writes:

The measurement instrument had to take care not to ask questions that directed subjects’ attention to the evaluation task as a ‘dialect thing’. Our choice of evaluative items in terms of *personality* traits, as well as the particular adjective pairs we chose to represent these traits, was based on experiences and results from our previous research in Denmark, which has allowed us to collect subconscious attitudes from a large number of audiences. (Italics ours)

We will demonstrate that Kristiansen’s insistence on personality traits (in order to hide the attitudinal object from the respondents) erroneously excludes *speech* traits from SE-based language attitude research. In the following section, we report a Free Response (FR) experiment to demonstrate that in addition to speaker conceptualizations, language attitudes in the Netherlands also contain *speech-related* perceptions of the aesthetics and functional appropriateness of the investigated varieties. Section 3 reports an SE experiment which demonstrates that the inclusion of speech-related scales in an SE experiment does not affect the nature and the structure of the attitudes elicited, so that there is no methodological impediment to including them in SE research. In section 4 we flesh out some of the theoretical and methodological consequences of our findings.

*The authors are indebted to Mieke Steegs for her assistance in the practical implementation of the experiments reported here.

2 Should Language Attitudes Contain Speech-Related Dimensions?

2.1 Background

In the SE technique pioneered by Lambert et al. (1960), listener-judges rate recorded samples of language varieties on a number of evaluative scales. On the resulting set of ratings, factor analysis is applied to detect the basic dimensions of the architecture of language attitudes (see Garrett 2005, Garrett et al. 2003, for an overview and evaluation). Because the standard practice in the SE tradition is to select one's rating scales from a number of standard studies (see Zahn and Hopper 1985 for an overview), many investigations have found accordingly that their ratings correlate into two or three basic dimensions, typically Speaker Status, Speaker Solidarity (social attractiveness), and sometimes also Speaker Integrity.

In the Dutch language area, Heijmer and Vonk (2002) reported an SE experiment which confirmed the Status- and Solidarity-based architecture of Dutch accent attitudes, but they also found that all regional accents obtained virtually identical low scores on the Status dimension, as well as comparably high scores on the Solidarity dimension. Likewise, Speelman and Impe (2007) conducted an SE experiment which returned similar Status, Solidarity and Integrity scores for Belgian Standard Dutch and substandard Dutch from Limburg (the eastern-most province of Flanders whose inhabitants and their accent are traditionally considered to be rural and unsophisticated).

In order to address the concern that it is the small number of attitude determinants (Status and Solidarity/Integrity) across which perceptions are compared which equalizes accent attitudes that may be quite different in reality, Grondelaers et al. (2010) proposed a richer measurement instrument containing 18 scales selected with a view to detecting attitudinal regularity beyond the Status, Solidarity, and Integrity dimensions. A stratified sample of listener-judges ($n=133$) rated four regional accents of Netherlandic Standard Dutch, each represented by 20 second samples of spontaneous non-edited speech from two speakers. As predicted, factor analysis returned a comparatively richer attitude architecture than in most other SE work: the analysis confirmed the basic dimensions Speaker Status and Speaker Integrity, but it also returned two dimensions pertaining to the accents *themselves*, namely Speech Status—how appropriate is an accent for formal interaction?—and Speech Euphony—which accents sound nice and which do not? While these data confirm the ubiquity of the Status and Attractiveness dimensions in language attitudes, they indicate, at the same time, that these dimensions need a *social* (with respect to speakers) as well as a *linguistic* (with respect to speech) application in order for researchers to fully appreciate the nature and structure of language attitudes.

While the latter in itself constitutes valid evidence in support of the view that SE experiments should go beyond personality traits, it would be interesting to find out whether a *direct* measurement technique which does not rely on scaled dimensions given a priori returns a similar attitude architecture. More specifically: if listener-judges are required to explicitly articulate their evaluations, will these evaluations reveal the same underlying dimensions as the ones underlying the SE data and, more importantly, to what extent do these perceptions pertain to the speakers and to their speech?

2.2 Method

A stratified sample ($n=116$) of native speakers of Netherlandic Dutch (47 male vs. 69 female; age range 13–63, with an average of 29.63; 52 from the South, 12 from the North, 15 from Randstad, 36 from the East) were asked to name the first three adjectives which came to mind when confronted with the labels of eight regional varieties of Netherlandic Dutch (Limburgs, Brabants, Gelders, Gronings, Achterhoeks, Twents, Fries, Hollands) and the substandard variety of Belgian Dutch (Vlaams). Participants were encouraged to articulate their evaluations in terms of adjectives (in order to tap into evaluative dimensions), and to do so as quickly as possible (in order to access spontaneous, unreflected perceptions as much as possible).

This free response (FR) experiment returned 3,089 tokens of 557 adjective types, 273 of which were hapaxes (49%). Table 1 lists the absolute frequencies ($n>1$) of the adjective types associated with the labels “Limburgs,” the low prestige variety spoken in the Southern-most rural province of Limburg, and “Hollands,” the high prestige variety spoken in the Western Randstad,

the urban concatenation of major cities such as Rotterdam, Amsterdam and Utrecht. These labels represent varieties which are in almost diametric opposition on any conceivable perceptual trait.

Dimension	Adjective	Translation	<i>Limburgs</i>	<i>Hollands</i>
<i>Speaker Status</i> 13,12	beperkt	limited	2	
	boers	boorish	9	
	dom	stupid	20	4
	dominant	dominant		2
	intelligent	intelligent		2
	naief	naïve	4	
	ordinair	vulgar, common		3
	simpel	simple	2	2
	slim	smart		5
	sloom	dull	2	
	stom	stupid	2	
	vlot	easy, flowing		2
	volks	vulgar		6
	werelds	worldly		3
	voortuitstrevend	progressive		2
<i>Speaker Attractiveness</i> 49,50	aardig	nice	3	2
	achterdochtig	untrusting	2	
	afstandelijk	uninvolved		7
	arrogant	arrogant		21
	asociaal	antisocial		13
	bekakt/kakkerig	posh		7
	besloten	private	2	
	bourgondisch	Burgundian	7	
	brutaal	brutal		4
	chauvinistisch	chauvinistic	2	
	direct	direct		19
	druk	(over)energetic	2	2
	éénkennig	shy	2	
	eerlijk	honest		5
	egocentrisch/egoïstisch	egocentric		3
	eigen	private	2	
	eigenwijs/eigenzinnig	opinionated	3	
	extravert	extroverted		2
	feestelijk	convivial, festive	3	
	fel	fierce, keen		2
	gastvrij	hospitable	6	
	geïsoleerd	isolated	2	
	gemoedelijk	genial	4	
	gesloten	reticent	4	
	gezellig	sociable	51	9
	grappig	funny	5	3
	indirect	indirect	3	
	individualistisch	individualistic		2
	irritant	irritant	6	
	joviaal	jovial		2

	lief	sweet	3	
	luidruchtig	loud		4
	nuchter	hard-headed		4
	open	open	4	
	opschepperig	boastful		2
	overdreven	exaggerated		7
	relaxed	relaxed	2	
	rooms	Roman Catholic	2	
	sociaal	social		2
	trots	proud	5	
	vervelend	boring, dull	2	
	vriendelijk	friendly	7	
	vrolijk	cheerful, lively	4	
	vrouwelijk	feminine	2	
	zachtaardig	gentle, mild	3	
	zuinig	stingy		3
<i>Speech Status</i> 7,55	buitenlands	foreign	4	
	duidelijk	clear, explicit		7
	duits	German	7	
	lokaal	local	2	
	moeilijk	difficult		2
	onverstaanbaar	unintelligible	6	
	reëel	real		2
	standaard	standard		3
<i>Speech Attractiveness</i> 11,73	verstaanbaar	intelligible		3
	vertrouwd	familiar	3	2
	langzaam	slow	2	
	lelijk	ugly	5	3
	mooi	beautiful	2	3
	plat	vulgar	5	5
	schel	piercing		2
	scherp	sharp		4
<i>Speaker Status/ Speech Status</i> 6,56	snel	fast		3
	smeuig	savory	2	
	zangerig/zingend	melodious	25	
	anders	different	2	
	beschaafd	civilized		4
	formeel	formal		2
	gewoon	ordinary		2
	goed	good		3
<i>Speaker Attractiveness/ Speech Attractiveness</i> 11,53	modern	modern		3
	multicultureel	multi-cultural		3
	netjes	neat, tidy		2
	normaal	normal		5
	stads/stedelijk	urban		10
	hard	hard		19
	leuk	nice	5	3
	schreeuwerig	shrieking		2
	traag	slow	3	
	warm	warm	7	

	zacht	soft	19	
<i>Total</i>			257	246

Table 1: Absolute frequency of adjective types ($n > 1$) returned for the regional variety labels "Limburgs" and "Hollands" in a Free Response task.

In addition to the different adjective types and their frequencies for the two language variety labels, Table 1 also distinguishes classifiers—in column 1—which group the adjectives in terms of the underlying attitude components they seem to instantiate. This classifying exercise “reverses” the design process which leads to scale selection in SE experiments. In the latter, the attitudinal dimensions the researcher wants to confirm determine the choice of (bi-polar or Likert-scaled) adjectives. In the present experiment, this causal relationship is reversed because we *infer* the underlying dimensions from the adjectives which reflect our participants’ spontaneous evaluations. If attitude architecture (as revealed in SE experiments) is determined by a finite number of underlying dimensions along which attitudes vary and change, then it is attractive to surmise that these dimensions will also surface in an FR task (where they determine the adjectives our participants name).

It is revealing to notice in this respect that the adjectives obtained in the FR experiment can straightforwardly¹ be classified into the four dimensions confirmed in the factor analysis in the SE experiment reported in Grondelaers et al. 2010, namely Speaker Status, Speaker Attractiveness, Speech Status, and Speech Attractiveness. While there is clearly room for finer-grained classification—because the Speaker Status adjectives can be further divided into Competence, Superiority, and Dynamism adjectives, and the Speaker Attractiveness class contains Integrity and Solidarity adjectives—it is obvious that both the SE and the FR experiment tap into highly similar underlying attitudes. Most crucial for the purpose of this paper is the observation that there are two classes of adjectives (totaling 19.28% of all the adjectives returned more than once) which clearly pertain to the varieties investigated rather than to their speakers: adjectives such as *clear*, *(un)intelligible*, *standard* can only pertain to the status of the varieties themselves, while *melodious* or *piercing* cannot but pertain to the sound aesthetics of Limburgs and Hollands. Observe also that the adjective *melodious*, which is the second-most frequently returned in Table 1 ($n=25$), is crucial for the perceptual characterization of the Limburg variety, whose melodious “whine” is among the most persistent stereotypes associated with Southern speech.

This experiment confirms in any case that speech evaluations are a substantial component of the attitudes towards regional varieties of Dutch, in addition to speaker/person traits. The next issue is then: to what extent does the inclusion of speech-related scales in an SE design affect the nature or quality of the attitudes investigated? Do language attitudes become more “conscious,” and hence less informative, when listener-judges can infer the attitudinal object—speech instead of its speakers—from the experimental task?

3 Do Language Attitudes Become More Conscious When Investigated with an SE Experiment Containing Speech Traits?

3.1 Listener-Judges and Speech Samples

The experiment was designed with the methodological purpose discussed in this paper, but also with an empirical purpose (which will be briefly outlined here because it affects the composition and stratification of the participant sample). The empirical purpose addressed the question of how inhabitants from the Belgian and Netherlandic provinces of Limburg—an area which originally

¹While some amount of indeterminacy and ad hoc decision-making is (inevitably) inherent to this type of exercise, the classification into the factor labels in column 1 is reassuringly straightforward in the majority of cases. The major problem concerns the fact that it is not always clear whether decontextualized adjectives such as *hard* or *modern* refer to Limburg or Holland speakers or speech, hence the overlapping categories at the bottom of Table 1.

belonged to Belgium before part of it became Dutch in 1839—perceive their own and each other's varieties of Dutch. To this effect, 392 secondary school pupils from eight schools² in Belgian Limburg (k=4) and Netherlandic Limburg (k=4) evaluated unedited 20 second samples of speech from two Belgian Limburg speakers, two Netherlandic Limburg speakers, one Belgian speaker with a central accent, one Belgian speaker with a peripheral accent, one Netherlandic speaker with a central accent, and one Netherlandic speaker with a peripheral accent.

3.2 Conditions and Scales

Rating scales were constructed in two conditions which had nine scales in common: six of the latter were adapted from earlier experiments, and three new scales (*insecure*, *passive* and *slow*) were added to investigate the hypothesis that Netherlandic Limburg is perceived as more *dynamic* than Belgian Limburg.

Six additional scales were constructed in two conditions which varied with respect to their serving to refer to speakers in Condition 1, but to speech in Condition 2:

Condition 1

Deze persoon **is geschikt als** nieuwslezer.
'This person qualifies as a news anchor.'
Deze persoon **is geschikt als** journalist.
'This person qualifies as a journalist.'
Deze persoon **is geschikt als** presentator.
'This person qualifies as a radio presenter.'

Deze persoon **is** traag.
'This person is slow.'
Deze persoon **is** koud.
'This person is cold.'
Deze persoon **is** zacht.
'This person is soft.'

Condition 2

Deze persoon **klinkt als** een nieuwslezer.
'This person sounds like a news anchor.'
Deze persoon **klinkt als** een journalist.
'This person sounds like a journalist.'
Deze persoon **klinkt als** een presentator.
'This person sounds like a radio presenter.'

Deze persoon **klinkt** traag.
'This person sounds slow(ly).'
Deze persoon **klinkt** koud.
'This person sounds cold(ly).'
Deze persoon **klinkt** zacht.
'This person sounds soft(ly).'

The first three scales were designed to instantiate the Speaker Status dimension in condition 1, and the Speech Status dimension in condition 2. The next three scales were designed in function of the Speaker Attractiveness dimension in condition 1 and the Speech Attractiveness dimension in condition 2. Recall that the second group of scales builds on adjectives (*slow*, *cold*, *soft*) which were found to be referentially ambiguous between speakers and speech when decontextualized. In this design, the wording forces either the speaker- or the speech-application.

3.3 Procedure and Instructions

Data were collected by trained student-assistants who had prepared two types of booklets in advance of the experiment (one for each condition), which contained eight response sheets for the eight experimental speakers, as well as a data sheet on which some demographic properties of the listener-judges were elicited. The two conditions were distributed alternatim among the students, who were told that their neighbors had different questionnaires so that it would be useless to cheat. The general instructions provided for all respondents were that they were participating in a study concerned with how people rate personalities on the basis of limited information (see Abrams and Hogg 1987).

²Secondary schools were preferred in this experiment for practical and logistic reasons (a sufficiently large sample of respondents who could take the experiment simultaneously, viz. in class). Schools were chosen in comparable locations in both provinces (two alongside the river Meuse—the natural border—one central, one peripheral). All respondents came from the three highest grades (4th to 6th) of the most advanced type of pre-university training in the Belgian and Dutch educational system. 204 listener-judges were male, 188 were female.

3.4 Results

A separate Principal Component Analysis (PCA) was carried out (SPSS, procedure Factor Analysis, factor selection criterion Eigenvalue >1, Varimax rotation) on the ratings in the two conditions. Tables 2 and 3 respectively contain the PCAs for condition 1 (speaker scales only) and condition 2 (speaker and speech scales):

	Speaker Attractiveness	Speaker Competence	Speaker Dynamism	Speaker Superiority
unfriendly	0,744	0,056	0,159	0,138
passive	0,223	0,287	0,701	0,180
dishonest	0,633	0,046	-0,024	0,190
stupid	0,113	0,211	0,207	0,808
lowly educated	0,038	0,232	0,188	0,826
insecure	0,029	0,147	0,547	0,476
asocial	0,653	0,085	0,362	0,115
dull	0,107	0,123	0,750	0,128
uncivilized	0,523	0,172	0,015	0,489
news anchor	0,132	0,794	0,234	0,251
slow	-0,056	-0,344	0,686	-0,096
journalist	0,136	0,791	0,250	0,243
cold	0,646	-0,138	-0,300	0,113
soft	-0,757	0,101	-0,085	-0,076
presenter	0,129	0,841	0,229	0,123

Table 2: PCA of ratings on the basis of **speaker and speech scales**; bold face for loadings >.4; 63.718 % variance explained.

	Speaker Attractiveness	Speech Status	Speaker Dynamism	Speaker Superiority
unfriendly	0,759	0,105	0,058	0,059
passive	0,398	0,271	0,609	0,183
dishonest	0,670	0,028	-0,093	0,223
stupid	0,243	0,208	0,217	0,742
lowly educated	0,049	0,103	0,122	0,848
insecure	0,282	0,217	0,462	0,329
asocial	0,751	0,049	0,188	0,158
dull	0,316	0,262	0,684	0,172
uncivilized	0,535	0,145	0,051	0,473
news anchor	0,058	0,829	0,083	0,196
slow	-0,129	-0,200	0,775	-0,119
journalist	0,056	0,836	0,174	0,143
cold	0,601	-0,017	-0,317	0,040
soft	0,340	0,060	0,606	-0,007
presenter	0,090	0,811	0,185	0,040

Table 3: PCA of ratings on the basis of **speaker scales** only; bold face for loadings >.4; 63.718 % variance explained.

It is revealing to note that the different wording of six scales had no effect whatsoever on the basic dimensions they correlate into: there is a remarkable parallelism between the factor solutions the

PCA suggests for the ratings in the two conditions. In both conditions, a four factor solution emerged with three identical components on which (almost) the same scales received virtually identical loadings. In addition to Speaker Attractiveness, there is evidence for a Speaker Dynamism dimension (as predicted) and a Speaker Superiority/Inferiority dimension. Interestingly, our transformation of Speaker Status into Speech Status scales in condition 2 had no discernible effect on the PCAs, both of which returned a separate factor on which the manipulated scales load. The only difference is in the interpretation: the factor which surfaces as a Speaker Status/Competence dimension in condition 1 (Table 2) becomes a Speech Status dimension in condition 2 (Table 3).

The effect of our manipulation of the referentially ambiguous adjectives *slow*, *soft*, and *cold* is more difficult to pinpoint. In condition 1, coldness and hardness load on the Speaker Attractiveness dimension, while slowness is considered to be an attribute of a non-dynamic personality. Condition 2 replicates these associations, with the difference that *speaking* softly has now also become an attribute of a non-dynamic rather than an unattractive personality. The fact that *slow*, *soft*, and *cold* do not correlate into a separate Speech Euphony dimension suggests that we have been unable to “redirect” the adjectives *slow*, *soft*, and *cold* to the speakers’ speech rather than to their personality. Observe that it is in any case difficult to replicate the Speech Euphony dimension in an SE experiment (no matter how easily and recurrently it can be confirmed in an FR task). Factor analysis in a related experiment (Grondelaers and Van Hout 2010) did not return the Euphony dimension either, although similar sound-inducing scales had been added. Admittedly, speech euphony was jeopardized in this experiment because speech stimuli were segmented in function of the implementation of a social variable (the experiment investigated the competition between regional accent and professional information as impression determinants). In the current experiment, a possible explanation for the absence of the Euphony dimension is the fact that Belgian Dutch samples and Belgian Dutch respondents were involved: prior research on Belgian Dutch (Grondelaers et al. 2009) has shown that sound quality is not central to Belgian perceptions of regional accent variation.

The major conclusion to be drawn from this experiment is that the presence of speech-related scales in an SE task does not affect the nature of the investigated attitudes: if attitudes returned in an SE experiment which only contains speaker traits are subconscious, then attitudes offered in an SE experiment which contains speaker *and* speech traits are subconscious as well. This finding counters Kristiansen’s (2009) suggestion that the consciousness or subconsciousness of the resulting perceptions correlates with the respondents’ awareness of the attitudinal object.

4 Conclusion and Consequences

On the basis of a Free Response experiment we have demonstrated, first, that attitudes in the Dutch language area pertain to speakers *and* to their speech, and that attitude studies should accordingly integrate scales pertaining to both. A follow-up Speaker Evaluation experiment confirmed that the inclusion of speech traits in a Speaker Evaluation experiment does *not* affect the nature and structure of the attitudes observed. The data reported counter Kristiansen’s (2009) claim that the inclusion in an SE experiment of scales which reveal the attitudinal object to the respondents may render the resulting attitudes more conscious and therefore less valuable.

In the course of the research, however, an even further-reaching conclusion surfaced: the FR and SE experiments appeared to return *the same* attitudinal content and structure. Why bother with an SE experiment then, when an (easier to implement) FR task seems to do the job just as well?

Although time and space restrictions preclude the extensive discussion this pivotal question merits, we will briefly sketch three reasons for carrying out both types of data collection in attitude research. Observe, to begin with, that the FR task requires robust and unambiguous labels which univocally identify a speech variety and its speakers to linguistically naive native speakers. Such generally shared labels are not always available, especially not in the case of new, developing varieties or when different social groups have developed their own sets of labels. It should be noted in this respect that the label *Hollands* which was used in Table 1 is problematic. Although it identifies the South- and North-Holland provinces in the western Netherlands, it is often used to refer to the whole of the Netherlandic territory (the song *Hup Holland Hup* ‘Go Holland Go’, which is the vocal trademark of Dutch soccer fans, bears testimony to this fact). In addition, it carries posh connotations which are by no means shared by all the inhabitants in the region. As a result, the

label *Hollands* invites contradictory associations, even by its own inhabitants, who find their fellow-Hollanders superior and sophisticated, and vulgar and brutal at the same time. In some cases, there is no label yet: Stroop (1998) has pointed linguists and laymen to the emergence of what he calls a new “variety” of Dutch, which is phonetically characterized by lowering of the diphthongs. While most of the Dutch will have become to some extent aware of this change, Stroop’s label *Poldernederlands* (Polderdutch) is only known to trained linguists. There is no point, therefore, in using this label in an FR task with lay speakers. In this situation, it is therefore more advantageous to design an SE experiment with unlabeled speech stimuli.

A second reason to continue both attitude measurement techniques is the fact that a preliminary FR experiment is of enormous heuristic value for the scale-selection phase in an SE experiment. A carefully prepared FR task will reveal the attitudinal dimensions to be included in the SE experiment (and the scales to be selected accordingly). We hope to have demonstrated that the FR task reported above uncovered all the attitudinal dimensions that were later confirmed in the SE experiment, and at a finer level of granularity too: recall that the adjectives in the Speaker Status category in Table 1 could be further subclassified into Superiority, Competence and Dynamism adjectives.

Third, and most importantly, SE experiments which build on unlabeled speech fragments can return perceptions to which respondents have no other (direct) access. Speech fragments, to begin with, can be carefully manipulated to contain varying sets of linguistic features and/or voice characteristics, while it is unclear which linguistic and non-linguistic triggers are activated by speech variety labels (see also Preston 2009); it is not unlikely, moreover, that the same label activates different triggers for different (groups of) listeners. More specifically, SE experiments can be designed to access changing attitudes or even new attitudes that are not articulated yet in the explicit evaluative repertoire of a speech community. A case in point is the changing status of Limburg-accented speech³ in The Netherlands which is confirmed in an SE experiment (Grondelaers et al. 2010) but not (or not yet) in the FR task reported here, which returns the low speech and speaker status stereotypes always associated with the Limburg area.

While the latter data confirm Kristiansen’s (2009) claim that it is the SE experiment which returns the more dynamic evaluations reflecting language (perception) change, we are reluctant to equate the difference between SE and FR with a difference in consciousness. If anything, spontaneous speech can tap “deeper” into underlying dimensions than naked labels: the evaluative structure revealed by spontaneous speech and language variety labels need not be radically different or function on different levels of consciousness.

References

- Abrams, Dominic, and Michael A. Hogg. 1987. Language attitudes, frames of reference, and social identity: A Scottish dimension. *Journal of Language and Social Psychology* 6:201–213.
- Garrett, Peter. 2005. Attitude measurements. In *Sociolinguistics. An international handbook of the science of language and society*, ed. U. Ammon, N. Dittmar, K. J. Mattheier and P. Trudgill, 1251–1260. Berlin: De Gruyter.
- Garrett, Peter, Nicholas Coupland, and Ann Williams. 2003. *Investigating Language Attitudes: Social Meanings of Dialect, Ethnicity and Performance*. Cardiff: University Of Wales Press.
- Grondelaers, Stefan, and Roeland van Hout. To appear. Is Standard Dutch with a regional accent standard or not? Evidence from native speaker attitudes. To appear in *Language Variation and Change* 22:1–19.
- Grondelaers, Stefan, Roeland van Hout, and Dirk Speelman. 2009. The robustness and stability of accent attitudes in Netherlandic and Belgian Standard Dutch. Paper presented at ICLaVE 5, University of Copenhagen.
- Grondelaers, Stefan, Roeland van Hout, and Mieke Steegs. 2010. Evaluating regional accent variation in Standard Dutch. *Journal of Language and Social Psychology* 29:101–116.
- Heijmer, Tessa, and Roos Vonk. 2002. Effecten van een regionaal accent op de beoordeling van de spreker. *Nederlands Tijdschrift voor de Psychologie* 57:108–113.

³There are unmistakable signs that Limburg-accented speech is losing some of its backwardness in the eyes of the Dutch.

- Impe, Leen, and Dirk Speelman. 2007. Vlamingen en hun (tussen)taal. Een attitudineel mixed guise onderzoek. *Handelingen van de Koninklijke Maatschappij voor Taal- en Letterkunde en Geschiedenis* 16:109–128.
- Kristiansen, Tore. 2009. The macro level social meaning of late modern Danish accents. *Acta Linguistica Hafniensia* 41:167–192.
- Lambert, Wallace E., Richard C. Hodgson, Robert C. Gardner, and Stanley Fillenbaum. 1960. Evaluative reactions to spoken languages. *Journal of Abnormal and Social Psychology* 66:44–51.
- Preston, Denis. 2009. Are you really smart (or stupid, or cute, or ugly, or cool)? Or do you just talk that way? In *Language Attitudes, Standardization and Language Change – Perspectives on Themes Raised by Tore Kristiansen on the Occasion of his 60th Birthday*, ed. M. Maegaard, F. Gregerson, P. Quist and J. N. Jørgensen, 105–29. Oslo: Novus Forlag.
- Stroop, Jan. 1998. *Poldernederlands; Waardoor het ABN verdwijnt*. Amsterdam: Bert Bakker.
- Zahn, Christopher J., and Robert Hopper. 1985. Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology* 4:113–123.

Centre for Language Studies
Radboud University of Nijmegen
Postbus 9103
6500 HD nijmegen
S.Grondeelaers@let.ru.nl
R.v.hout@let.ru.nl